

METHOD AND APPARATUS FOR AUDIO NAVIGATION OF AN INFORMATION APPLIANCE

FIELD OF THE INVENTION

[0001] The present invention relates, generally, to Internet-capable appliances and, more specifically, to methods and apparatus for configuring such appliances for audio navigation.

BACKGROUND OF THE INVENTION

[0002] Electronic Program Guide (EPG) is a favorite channel on television because it helps navigate the user through a myriad of program choices. EPG, however, cannot be used by visually impaired persons because of the graphics-rich user interface. The many subliminal visual cues available to sighted users are absent for blind/visually impaired users. Visual information is not presented in an understandable format to the visually impaired, nor is data rearranged to suit an accessibility mode for the visually impaired.

[0003] Embedded text to speech (TTS) algorithms have been demonstrated in appliances to convert text-based EPG to audio-enabled EPG. These appliances are expensive, however, since a good quality TTS synthesizer is required in each appliance. Large storage capacity is also required to accommodate a TTS synthesizer.

[0004] A need exists, therefore, to provide an audio enabled system using an information appliance that is compatible with a visually impaired user, and does not require an expensive internal TTS synthesizer.

SUMMARY OF THE INVENTION

[0005] To meet this and other needs, and in view of its purposes, the present invention includes a method of providing information using an information appliance coupled to a network. The method includes storing text files in a database at a remote location and converting, at the remote location, the text files into speech files. The method also includes requesting a portion of the speech files. The portion of the speech files requested are downloaded to the information appliance and presented through an audio speaker. The speech files may include audio of electronic program guide (EPG) information, weather information, news information or other information.

[0006] The method may include downloading the speech files in response to a specific request, or downloading the speech files at periodic time intervals. The speech files may be stored or buffered in a memory device of the information appliance and later presented, through the audio speaker, in response to a request.

[0007] In another embodiment, the method includes converting the text files into speech files at the remote location using an English text-to-speech (TTS) synthesizer, a Spanish TTS synthesizer, or another language synthesizer. A voice personality from a list of multiple voice personalities may also be selected. In response to the selection, the method converts the text files into speech files using the selected voice personality.

[0008] It is to be understood that both the foregoing general description and the following detailed description are exemplary, but are not restrictive, of the invention.

BRIEF DESCRIPTION OF THE DRAWING

[0009] The invention is best understood from the following detailed description when read in connection with the accompanying drawings. Included in the drawings are the following figures:

[0010] FIG. 1 is an overview of an audio-enabled data service system according to an embodiment of the present invention;

[0011] FIG. 2 is an exemplary embodiment of an information appliance;

[0012] FIG. 3 is a basic workflow diagram illustrating steps involved in a typical operation executed via interfacing software according to an embodiment of the present invention;

[0013] FIG. 4 illustrates various options that may be selected by a user during the operation diagrammed in FIG. 3; and

[0014] FIG. 5 illustrates steps involved in navigating through an electronic program guide when the user selects a search option shown in FIG. 4.

DETAILED DESCRIPTION OF THE INVENTION

[0015] FIG. 1 is an overview of an audio-enabled data service system, generally designated by numeral 10. In the embodiment shown, audio-enabled data service system 10 includes text-to-speech (TTS) application server 20 communicatively coupled to integrated television 26 by way of Internet 24. Integrated television 26 includes information appliance 28 and television 30.

[0016] As will be explained, a user wishing to access TTS application server 20 may activate a setup procedure in information appliance 28 which then dials server 20. The user may call, or the appliance may automatically dial after obtaining permission from the user, a specific dial-up number provided to the user. The server may be accessed via a telephone connection established by a Service Control Point (SCP) located in a telephone network, such as Publicly-Switched Telephone Network (PSTN), wireless network or cableless network (not shown). In many cases, the user of information appliance 28 needs an Internet Service Provider (ISP) (not shown) to complete the connection, via the Internet, between information appliance 28 and server 20.

[0017] It is apparent to one skilled in the art that Internet 24 may be of another type of data network, such as an Intranet, private Local Area Network (LAN), Wide Area Network (WAN), and so on.

[0018] Having connected to TTS application server 20, interfacing software (not shown) in the server may recognize information appliance 28 by telephone number recognition via destination number identification service (DNIS) and automatic number identification (ANI). By recognizing information appliance 28, the server may select appropriate set-up routines to deal with the specific information appliance.

[0019] TTS application server 20 may include a large repository, which may be internal or separate from the server. Shown separate from server 20 in Fig. 1, the repository may include electronic program guide (EPG) database 12, weather database 14 and news database 16. As will be appreciated, additional databases containing other types of information may also be included, for example, a sports database.

[0020] In the embodiment shown, EPG information, weather information, and news information are stored as text. A text-to-speech (TTS) synthesizer is used to convert the text to speech (audio). A high quality text-to-speech software program may be resident in server 20, with versions to support multiple languages. As shown in FIG. 1, server 20 includes English TTS program 18 and Spanish TTS program 22.

[0021] When the user powers up the appliance for the first time, set-up information including software and protocol drivers may be delivered to information appliance 28 via the dial-up connection. In some cases, server 20 may communicate directly to a counterpart at the ISP and open an account for the appliance.

[0022] A resident audio program may prompt the user to select between text navigation or speech navigation. A normally sighted user may select text-navigation; a visually impaired user, on the other hand, may select audio-navigation. If the user selects audio-navigation, the resident program may provide a choice of different voices, including celebrity voices in various languages. A speech file may be

downloaded from the server to the appliance, and stored or buffered in the appliance for later, or immediate presentation to the user.

[0023] If the user selects text-navigation, text data may be downloaded from the server to the appliance. The text data may be stored in the appliance and later, or immediately displayed on television 30. Alternatively, a combination of text-navigation and audio-navigation may be selected by the user, in which case text data may be displayed on the television screen and audio data may be heard through audio speakers.

[0024] The files (speech, text or both) may be presented to the user as choices for easy navigation. When the user selects a choice, details of the choice may be presented. The user may also select, interrupt, or skip data by using a remote control. Navigation may be enriched by adding graphics to the audio and text data.

[0025] An exemplary embodiment of an information appliance is shown in FIG. 2 and is generally designated by the numeral 50. It will be understood that an information appliance may be a laptop, a desktop computer, a set-top box (STB), and the like, all of which are Internet-capable and are, therefore, Internet appliances. Exemplary information appliance 50 includes modem 60 connected or attached to telephone lines 66 for accessing the Internet via an ISP. Different types of data, including audio and text data, may be exchanged between information appliance 50 and TTS application server 20. The data exchanged may also include user identification, and preferences for downloading data from the server. The data may be formatted according to an application layer protocol having frame formats for telephone functions. These may include communications protocol hierarchy with Application Program Interface (API), Point-to-Point Protocol (PPP), and High-level Data Link Control (HDLC) layers for telephony applications.

[0026] It will be appreciated that although information appliance 50 is shown connected to telephone lines 66, it may be connected to a digital subscriber line (DSL), a twisted-pair cable, an integrated service digital network (ISDN) link, or any other link, wired or wireless, that supports packet switched communications,

including Internet Protocol (IP)/Transmission Control Protocol (TCP) communications using an Ethernet.

[0027] Information appliance 50 includes output devices, such as television 68 for displaying standard definition video and listening of audio through internal speakers. Stereo audio speakers 70, which are separate from television 68 may also be included. An input device, such as IR receiver 64, may be included for receiving control commands from user remote control 72.

[0028] Information appliance 50 includes processor 62 coupled by way of bus 54 to storage 52, digital converters 56 and graphics engine 58. Bus 54 collectively represents all of the communication lines that connect the numerous internal modules of the information appliance. Although not shown, a variety of bus controllers may be used to control the operation of the bus.

[0029] One embodiment of storage 52 stores application programs for performing various tasks, such as manipulating text, numbers and/or graphics, and manipulating audio (speech) received from telephone lines 66. Storage 52 also stores an operating system (OS) which serves as the foundation on which application programs operate and control the allocation of hardware and software resources (such as memory, processor, storage space, peripheral devices, drivers, etc.). Storage 52 also stores driver programs which provide instruction sets necessary for operating or controlling particular devices, such as digital converter 56, graphics engine 58 and modem 60.

[0030] An embodiment of storage 52 includes a read and write memory (e.g., RAM). This memory stores data and program instructions for execution by processor 62. Also included is a read-only memory (ROM) for storing static information and instructions for the processor. Another embodiment of storage 52 includes a mass data storage device, such as a magnetic or optical disk and its corresponding disk drive.

[0031] It will be appreciated that processor 62 may be several dedicated processors or one general purpose processor providing I/O engines for all the I/O functions (such as communication control, signal formatting, audio and graphics

processing, compression or decompression, filtering, audio-visual frame synchronization, etc.). Processor 62 may also include an application specific integrated circuit (ASIC) I/O engine for some of the I/O functions.

[0032] Digital converters 56, shown in Fig. 2, receive baseband video and audio signals (tuner not shown) from a broadcasting television station, and provide digital audio and digital video to processor 62 for formatting and synchronization. Prior to sending data to television 68 and speakers 70, processor 62 may encode audio-visual data in a unique format for presentation and listening (e.g., an NTSC, SDTV, or HDTV format for television).

[0033] Files stored as text and speech at server 20 (FIG. 1) may be received at information appliance 50. Speech (audio) may be received in various formats, such as AAC, MP3, WAV, etc, and may be compressed to save bandwidth. Resources for processing the data (text and speech) may be provided by processor 62, and may include resources for Internet access (Internet application programs), resources for producing a compatible display of text and graphics on television monitor 68, resources for implementing synchronized audio, and resources for control of information through a remote keypad control, such as infrared remote control 72.

[0034] FIG. 3 is a basic workflow diagram illustrating steps involved in a typical operation executed via interfacing software according to an embodiment of the present invention. The method shown in FIG. 3, generally designated by reference numeral 80, is described below.

[0035] A user plugs in a specific appliance, such as information appliance 50 of FIG. 2, and insures that all hardware connections are correct (step 81). The user calls or the appliance dials, after obtaining user permission, a specific dial-up number. The appliance is then connected to TTS application server 20. After confirming identity, a set-up application is launched to access protocol information and network drivers.

[0036] After the appliance is successfully set-up, a clear-for-operation signal may be issued for the user to begin using the appliance. In step 82, a voice may

prompt the user to “select configuration”. The user may, for example, first hear “visual mode?”. Secondly, the user may hear “audio mode?”. Thirdly, the user may hear “both, visual and audio modes?”. The user may select audio (step 83), corresponding to “audio mode?”; text/graphics only (step 85), corresponding to “visual mode?”; or audio and text/graphics (step 84), corresponding to “both, visual and audio modes?”.

[0037] Using remote control 72 (FIG. 2) the first, second, or third configuration may be selected by pressing any key immediately after hearing the specific configuration announced. The selected configuration may be announced again, thereby confirming user selection.

[0038] A voice may prompt the user to select from a list of different languages (step 86). For example, the user may first hear “English?”. Secondly, the user may hear “Spanish”? and so on. Again, using the remote control, the user may select the first (English), second (Spanish), or another language by pressing any key immediately after hearing the specific language announced. The selected language may be announced again, thereby confirming user selection.

[0039] A voice may prompt the user to select from a list of different voices (step 87). For example, the user may first hear a male voice saying “Mel Gibson?”. Secondly, the user may hear a female voice saying “Marilyn Monroe?”. Thirdly, the user may hear a cartoon voice saying “Donald Duck?”. Again, using the remote control, the user may select a voice by pressing any key immediately after hearing the specific voice announced. The selected voice may be announced again, thereby confirming user selection.

[0040] It will be appreciated that the steps described above may vary widely according to desired implementation. For example, if the user selects the text/graphics only configuration in step 85, language selection (step 86) and voice selection (step 87) may be skipped.

[0041] Having selected configuration, language and voice, the method enters step 88 to select download frequency. Files from the server may be periodically

downloaded every night at a preset time, or upon a specific request by the user. For example, if the appliance is a set-top box (STB) and is Internet-ready, the STB may periodically download audio and text files every night at midnight containing electronic program guide (EPG) information of scheduled television programs for the next day. Alternatively, the STB may download audio-enabled EPG files upon a specific request from the user. The downloaded files may be stored or temporarily buffered in the appliance. In this manner, a visually impaired user may enjoy audio-enabled EPG.

[0042] When the EPG or Guide button (for example) is selected on the remote control (step 89), the method enters step 90 allowing the user to navigate through the downloaded files using the remote control. As shown in FIG. 4, once inside the EPG, one of several options for navigating through EPG content may be selected. The options may include current time (step 92), date (step 94) and search (step 96). The options may be presented to the user in sequence, with pauses between sequences. For example, the user may first hear "current time?". The user may select the current time option by pressing any key on the remote control. The audio may then announce the following: 10:00 p.m. (brief pause), Channel 2-CNN Larry King Live (brief pause), Channel 3-Fox Baseball, Red Sox vs. Yankees (brief pause), Channel 4-(and so on). Accordingly, the audio may sequence through every program offered at 10:00 p.m. Next, the audio may sequence through every program offered at 10:30 p.m. (and so on).

[0043] The user may interrupt the sequence at any time by simply pressing an arrow key (for example) on the remote control. With no interruption from the user, the STB may continue announcing in sequence all the viewing possibilities until the list of offering is complete, wrapping from 10:00 p.m. to 10:30 p.m., then to 11:00 p.m., etc. Upon pressing an up-arrow key, the user may command the STB to interrupt the audio output. Upon pressing the up-arrow key again, the STB may be commanded to resume the audio output, picking up at the place of interruption.

[0044] The user may command the audio output to skip and begin at the next time slot (for example 10:30 p.m., the next major table) by pressing the up-arrow key twice in quick succession. The user may command the audio output to begin at the

next day by pressing the up-arrow key three times in quick succession. After a quick pause, the voice may continue announcing the list of offerings available at that date, time and channel.

[0045] The user may command the audio output to begin at a previous time slot or a previous date by pressing the down-arrow key twice in quick succession or three times in quick succession, respectively.

[0046] Returning to FIG. 4, the user may hear “date?” after first hearing “current time?”. The user may select the date option in step 94, by pressing any key on the remote control. The audio may then begin announcing the viewing possibilities starting at a specific date and time. For example, the audio output may announce the following: October 1, 10:00 p.m. (brief pause), Channel 2-CNN Larry King Live (brief pause), Channel 3-movie, Dracula Meets Jerry Springer (brief pause), Channel 4-(and so on). The user may continue navigating through EPG content in a manner similar to that described for the current time option.

[0047] It will be appreciated that if a sighted user and a visually impaired user are both using the EPG presentation, the preferred method is to select both the audio and text/graphics configuration in step 84 (FIG. 3). In one embodiment, the appliance may default to the audio and text/graphics configuration, if the user does not select any of the available configurations. In another embodiment, the appliance may store the selected configuration, so that the user will not need to select the same configuration again.

[0048] When the audio and text/graphics configuration is selected, server 20 may transmit the front page of the EPG for display on the television screen. Server 20 may also transmit the audio files, corresponding to the text on the page, for listening. These files may be transmitted serially for storage in the STB, and then played-back as the user is navigating the EPG. Alternatively, the files may be transmitted from the server, upon request by the STB, while the user is navigating the EPG.

[0049] In an embodiment of the invention, a sighted user may navigate the EPG text displayed on the screen. When the user focuses on a specific grid of the EPG, the audio portion corresponding to the specific grid may then be announced by voice. When the user focuses on another grid, the voice may announce the text (or legend) corresponding to the newly focused grid. For example, date/channel/time/legend audio files for a specific grid may be downloaded from the server and announced. In this manner, the sighted user and the visually impaired user may enjoy navigating the EPG together.

[0050] When the visually impaired user is navigating the EPG by himself, audio files of channel, date and time may be downloaded once for the entire EPG page displayed on the screen. Legends in each specific grid, however, may be downloaded only when the user stops or focuses on a specific grid. In this manner, when the user navigates, the STB may announce the position of the focus point, in terms of channel number, date and time. When the user focuses on a specific grid, the STB may announce the details on the specific grid.

[0051] It will be appreciated that files downloaded from the server may be selectively discarded from the STB. For example, when the audio storage or audio buffer is full, files may be discarded; when the program is finished, files may be discarded.

[0052] Completing the description of FIG. 4, a user may select the search option in step 96. If a visually impaired user selects the search option (as identified by selecting the audio-only configuration in step 83 of FIG. 3), the navigation process (generally designated by numeral 90 in FIG. 5) branches to step 101. The STB may sequentially announce available search categories, for example sports, movies, situation comedies, serial dramas, etc. In step 103, the user may listen to available search categories and in step 105, the user may select a category. Since a user may wish to hear all the available search categories before selecting the best choice, the STB may sequence though the available categories by announcing the choices more than once (shown as feedback from step 105 to step 101). As the desired category is again announced, the user may select the category by pressing any key on the remote control.

[0053] If a visually impaired user and a normally sighted user are both available for the search mode, navigation process 90 may branch to step 102. The sighted user may type a keyword, such as "sports" in step 102. As the keyword is typed on the remote control, the STB may announce each key typed. In step 104, the STB may return with the best matching results on the television screen and announce the same through the speakers. The user may then select the best category in step 106.

[0054] After selecting the desired choice or category, the STB may announce in step 107 the channel, date, time and legend. The user may select the announced channel, in step 108, or may sequence to the next listing.

[0055] Having described a visually impaired user listening to audio of EPG information, it will be appreciated that another embodiment of the invention includes a sighted user listening to an audio menu while driving a car. For example, the user may navigate through a news menu, weather menu, or sports menu while listening to audio information downloaded from a TTS server to an Internet appliance in the car.

[0056] It will be appreciated that the invention uses good quality TTS speech software at the server end. In this manner, cost of an information appliance is much lower since a TTS synthesizer need not be installed in the information appliance.

[0057] Although illustrated and described herein with reference to certain specific embodiments, the present invention is nevertheless not intended to be limited to the details shown. Rather, various modifications may be made in the details within the scope and range of equivalents of the claims and without departing from the spirit of the invention. It will be understood, for example, that the same concept may be extended beyond EPG to include other data services, such as weather, news, sports, etc.